VARIABLE BIT RATE CONTROL WITH TRELLIS DIAGRAM APPROXIMATION

Kei Kikuiri, Nobuhiko Naka, and Tomoyuki Ohya

Multimedia Laboratories, NTT DoCoMo, Inc.
3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8526, Japan
kikuiri@spg.yrp.nttdocomo.co.jp

Abstract

In this paper, we present a variable bit rate control method for speech/audio coding, under the constraint that the total bit rate of a super-frame to be a constant. The proposed method uses a trellis diagram for optimizing the overall quality of the super-frame. In order to reduce the computational complexity, the trellis diagram uses approximation by ignoring the encoder memory state between different paths. Simulations on the AMR Wideband show that the proposed variable bit rate control achieves up to 4.3 dB improvements to the constant rate coding in the perceptual weighted SNR.

1. Introduction

As multimedia applications for mobile communications such as video and audio streaming are becoming popular, the demand for sound quality is getting higher. Since these applications are one-way and not real-time, they have more relaxed delay requirements than conversational applications. This allows the coding parameters to be optimized since longer coding intervals are possible. As current cellular terminals already support multiple codecs, sound quality can be improved by using more than one coding schemes as well as coding bit rates and frame lengths.

Variable bit rate control schemes for speech coding are categorized into open-loop schemes with analytical methods and closed-loop schemes like Analysis-by-Synthesis. Open-loop schemes have lower computational complexity than closed-loop schemes and are considered to be more suitable for real-time applications. Most open-loop schemes utilize local speech features, for example voice detection [3][4][5], signal power [6], spectral entropy [7] and multi-level phonetic classification [8][9]. However, closed-loop schemes are more attractive since they can develop more optimal bit rates if the appropriate evaluation criterion is used.

A good example of a typical closed-loop scheme is FS-CELP (Finite State CELP). It selects the lowest bit rate which achieves predetermined weighted Signal-to-Noise Ratio (WSNR). Eriksson et al. proposed a scheme that uses cost function unified segmental SNR and bit rate [11]. Cellario et al. developed a hybrid open-loop and closed loop scheme consisting open-loop voicing decision and close-loop bit rate selection with perceptual weighted distortion power[12]. All of these schemes were applied to conversational applications.

This paper proposes a closed-loop variable bit rate control method using perceptual weighted distortion power without phonetic classification. With allowing longer coding delay, the method optimizes the total quality of a super-frame by controlling the bit rates of each frame, under the constraint of constant bit rate for the super-frame. It uses an approximated trellis diagram that represents the transition of averaged bit rate in a super-frame. This paper introduces simulation results for the AMR Wideband (AMR-WB) encoder standardized by 3GPP as well as by ITU-T G.722.2 [2]. The simulation results include the fundamental performance and the performance under various super-frame lengths in comparison with constant rate coding and exhaustive search.

2. Variable bit rate control with trellis diagram approximation

2.1. Super-frame structure

This paper assumes that the input signal is encoded so that each super-frame has the same bit rate. Fig. 1 shows that each super-frame consists of N frames. The super-frame k is encoded at the constant bit rate R, where

$$R = \frac{1}{N}\sum_{n=0}^{N-1} r_{k,n} \quad , \qquad (1)$$

and $r_{k,n}$ (n = 0, 1, …, N-1) is the bit rate of the frame n in the super-frame k. We locate the optimum set of $r_{k,n}$ that satisfies formula (1) while optimizing the quality of the decoded signal within the super-frame. This paper uses perceptual weighted distortion power as the criterion for quality optimization. The perceptual weighted distortion power $WD(n)$ of the frame n is defined as

$$\begin{aligned} WD(n) &= \sum_{t=0}^{T-1} wd(t)^2 \\ &= \sum_{t=0}^{T-1} \left(ws(t) - w\hat{s}(t)\right)^2 \end{aligned} \quad , \qquad (2)$$

where T is the length of the frame, $wd(t)$ is the perceptual weighted distortion, and $ws(t)$ and $w\hat{s}(t)$ are the perceptual weighted input signal and the perceptual weighted synthesized signal, respectively.

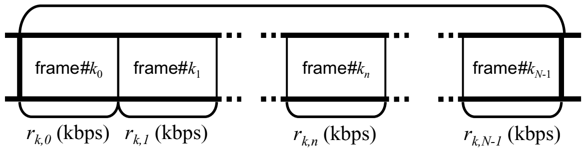


Fig.1. Super-frame structure.

2.2. Trellis diagram approximation

This paper proposes a variable bit rate control method that uses a trellis diagram to represent the transition of averaged bit rate in each super-frame. Fig. 2 shows the trellis diagram for the parameter set shown in Table 1. In this figure, the nodes denote the averaged bit rates after coding the corresponding frames. The branches and the paths denote the coded bit rates and the set of bit rates, respectively. This paper uses Viterbi algorithm with the metric of the perceptual weighted distortion power to locate the best path (the set of bit rates from the first to the last frame). Since the proposed method determines the best path super-frame by super-frame, the algorithmic delay is one super-frame.

An important point is that this trellis diagram is approximated so that the paths, whose averaged bit rates are equal, degenerate to one node. Actually, the memories in the encoder are different among the bit rate sets. For example, when we accurately illustrate the trellis diagram in Fig. 2, the node for bit rate 19.85 kbps at frame 3 is as described by Fig.3 (a). In comparison, the approximated trellis diagram (Fig.3 (b)) has fewer branches. Therefore, using the approximated trellis diagram reduces the number of encoding procedures.

Without the approximation technique, the trellis diagram is equivalent to exhaustive search. In the case of Table 1, the number of possible sets of $r_{k,n}$ is 75. In order to calculate the sum of the perceptual weighted distortion power for each $r_{k,n}$ set, the exhaustive search method needs to process the encoding procedure of 450 frames, because one super-frame consists of 6 frames in this case. This imposes excessive computational costs. On the other hand, the approximated trellis diagram has 53 branches. Consequently, the proposed method processes the encoding procedures of 53 frames to identify the best path. This represents an 88% reduction in computational complexity.

Due to this approximation, however, the best path is not strictly optimal. The next section uses computer simulations to evaluate the impact of the approximation.

Table 1. Example of parameters.

Bit rate R (kbps)	19.85
Super-frame length N (frame)	6
Selectable bit rates at each frame $r_{k,n}$ (kbps)	23.05 19.85 18.25

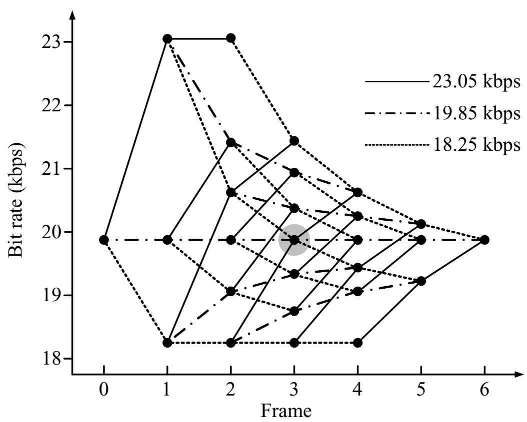
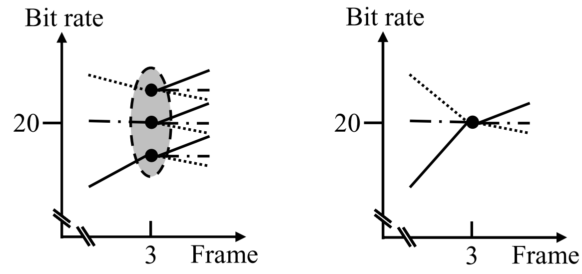


Fig.2. Trellis diagram.



(a) Without approximation. (b) With approximation.

Fig. 3. Example of trellis diagram with/without approximation.

3. Simulation results

3.1. Fundamental performance

The proposed method was implemented on AMR-WB with the parameters shown in Table 2. Since the method needs no modification of AMR-WB bit streams, output bit streams can be decoded by existing AMR-WB decoders. The quality of the method was compared with that of the constant rate coding (the bit rate of each frame is the same as that of the super-frame) and the exhaustive search method.

Both modes used the same trellis diagrams as shown in Fig.2. In mode 2, however, the dashed lines and the dotted lines correspond to 18.25 kbps and 15.85 kbps, respectively. The criterion to optimize the bit rate sets is the perceptual weighted distortion power calculated at 12.8 kHz sampling frequency; this is already used in the AMR-WB encoder.

The performance of proposed method is evaluated by the segmental WSNR, which is calculated at 12.8 kHz sampling frequency. The segmental WSNR $WSNR_{seg}$ is defined as

$$WSNR_{seg} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{l=0}^{L-1} ws(l)^2}{\sum_{l=0}^{L-1} wd(l)^2}, \quad (3)$$

where M is the number of segments and L (6 frames in this simulation) is the length of the segment.

Table 2. Parameters for the computer simulation.

	Mode 1	Mode 2
Bit rate R (kbps)	18.25	19.85
Super-frame length N (frame)	6	
Selectable bit rates at each frame $r_{k,n}$ (kbps)	23.05 18.25 15.85	23.05 19.85 18.25
Materials	Japanese speech, Japanese speech with BGM, piano, instrumental pop	

Fig. 4 and 5 show the segmental WSNR for each mode. More comprehensive results were reported in [13]. They show that the proposed method performs significantly better than constant rate coding, especially for speech signals. The maximum improvement is 4.3 dB. It offers a smaller improvement for the music signals, but this appears to be due to the poor capability of AMR-WB in handling music signals.

In comparison to the exhaustive search method, the segmental WSNR of the proposed method is degraded by only 0.5 dB even in the worst case. However, the computational complexity is reduced by 88% as mentioned in section 2.2. It can be said that the approximation of the trellis diagram significantly reduces the computational complexity at the cost of an insignificant degradation in decoded signal quality.

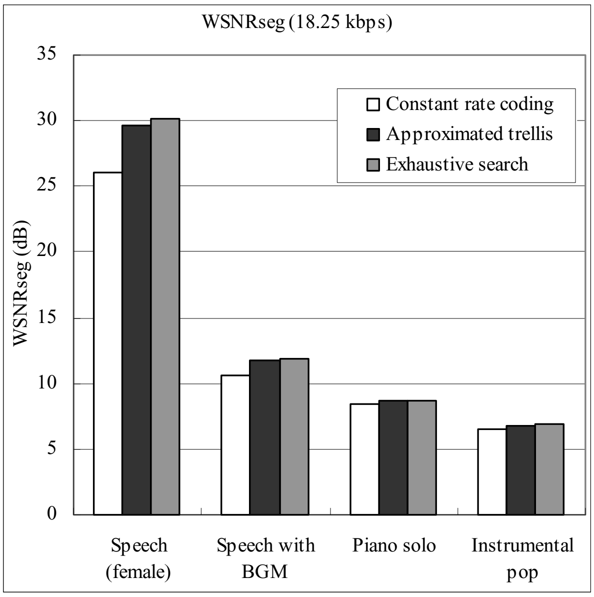


Fig. 4. Segmental WSNR of mode 1.

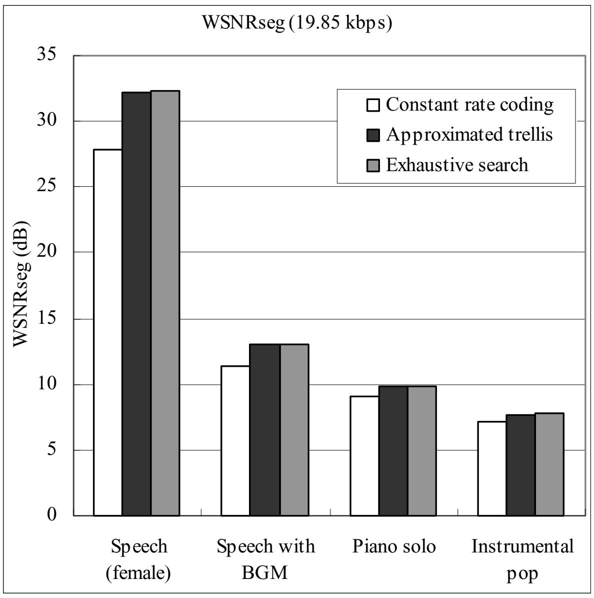


Fig. 5. Segmental WSNR of mode 2.

3.2. Performance under various super-frame lengths

This section shows the performance improvement from constant rate coding under the various super-frame lengths. The simulation in this section was also implemented on AMR-WB. The parameters and the trellis diagram are shown in Table 3 and Fig. 6, respectively. The number of selectable bit rates is reduced to 2 in order to make the simulation easier. The simulation was run on speech content and speech with BGM, for which the significant improvements were found in the previous section.

Table 3. Parameters for evaluating the performance improvement under various super-frame lengths.

Bit rate R (kbps)	19.85
Super-frame length N (frame)	3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36
Selectable bit rates at each frame $r_{k,n}$ (kbps)	23.05 18.25
Materials	Japanese speech (female and male), speech with BGM

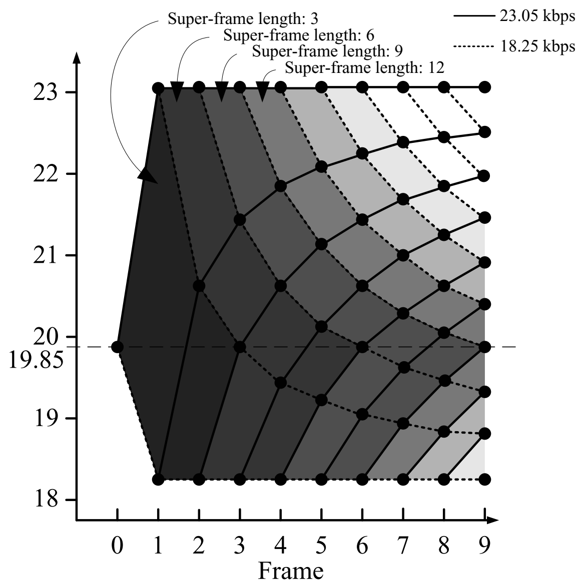


Fig. 6. Trellis diagram under various super-frame lengths.

Fig. 7 shows the Segmental WSNR improvement in comparison with the constant rate coding. The solid lines and the dotted lines denote the proposed method and the exhaustive search, respectively. The result of the exhaustive search method is obtained only up to the 12 frames super-frame due to the explosion of the simulation time. It is found that the proposed method offers the larger improvement with the longer super-frame length. However, the improvement is going to saturate around the 24 frames super-frame in this simulation.

Fig. 8 shows the computational complexity ratio to the constant rate coding. The proposed method requires more complexity for longer super-frame case. However, with the trellis diagram approximation, the proposed method drastically reduces the computational complexity compared to the exhaustive search method, while keeping the degradation from the optimal case small in the segmental WSNR as shown in Fig. 7. When the super-frame length is 12 frames, for instance, the proposed method performs one hundred times faster than the exhaustive search method and the degradation of the segmental WSNR is only 0.028 dB in the worst case. The advantage of the trellis diagram approximation becomes greater as the super-frame length extends.

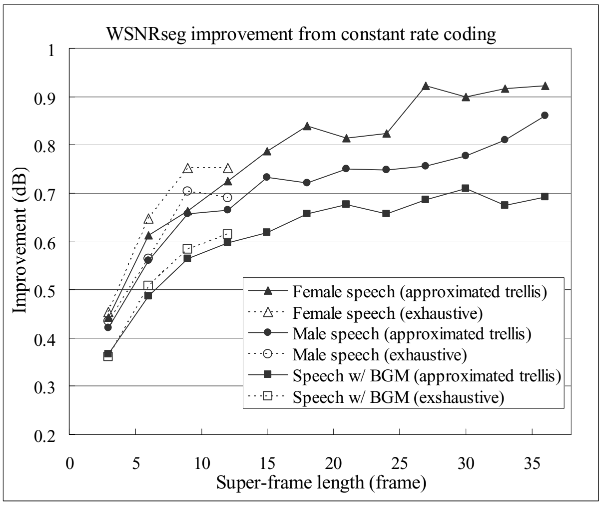


Fig. 7. Segmental WSNR improvement of various super-frame lengths.

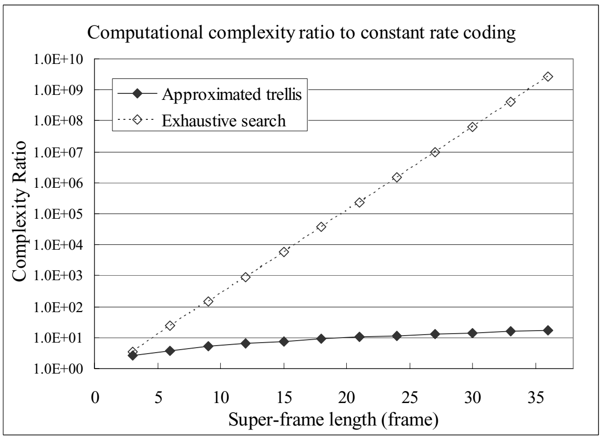


Fig. 8. Computational complexity ratio.

4. Conclusion

This paper proposed a variable bit rate control method that utilizes an approximated trellis diagram to optimize the bit rate allocation for frames in a super-frame; the assumption is that each super frame has the constant bit rate. The trellis diagram approximation is to ignore the difference of the memories in the encoder among the modes. Computer simulations of an AMR-WB implemented confirmed that the method offers significantly better performance than constant rate coding and achieved a maximum of 4.3 dB improvement in segmental WSNR. The segmental WSNR improvement is larger in case of longer super-frame lengths. While the longer super-frame requires more computational costs, the trellis diagram approximation reduces drastically the computational costs from the exhaustive search method and the quality degradation is insignificant. In addition, the reduction of the computational complexity becomes larger as the super-frame length gets longer. It can be applied to other codecs and also supports variable frame length control and variable coding scheme control.

5. References

[1] 3GPP TS 26.190, "AMR Wideband Speech Codec; Transcoding Functions," Feb. 2002.

[2] ITU-T G.722.2, "Wideband Coding of Speech at around 16 kbit/s using Adaptive Multi-rate Wideband (AMR-WB)," Jan. 2002.

[3] W. B. Kleijn, "Continuous Representations in Linear Predictive Coding," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process*., vol. 1, pp.201-204, 1991.

[4] Y. Shoham, "High-quality Speech Coding at 2.4 to 4.0 kbps based on Time – frequency Interpolation," *Proc. IEEE Int. Conf. Accoust., Speech*, Sig. Process., vol. 2, pp. 167-170, 1993.

[5] J. –M. Muller and B. Wachter, "A Codec Candidate for the GSM Half Rate Speech channel," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 1, pp. 257-260, 1994.

[6] A. DeJaco, W. Gardner, P. Jacobs and C. Lee, "QCELP: the North American CDMA Digital Cellular Variable Rate Speech Coding Standard," *Proc. IEEE Workshop on Speech Coding for Telecom.*, pp. 5-6, 1993.

[7] S. McClellan and J. D. Gibson, "Variable-Rate CELP based on Subband Flatness," *IEEE Trans. Speech and Audio Process.*, vol. 5, No. 2, pp. 120-130, Mar. 1997.

[8] E. Paksoy, K. Srinivassan and A. Gersho, "Variable Rate Speech Coding with Phonetic Segmentation," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 2, pp. 155-158, 1993.

[9] A. Das and A. Gersho, "A Variable-Rate Natural-Quality Parametric Speech Coder," *Proc. IEEE Int. Conf. on Commun.*, vol. 1, pp. 216-220, 1994.

[10] S. V. Vaseghi, "Finite State CELP for Variable Rate Speech Coding," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, pp.37-40, 1990.

[11] T. Eriksson and J. Sjoberg, "Evolution of Variable Rate Speech Coders," *Proc. IEEE Workshop on Speech Coding for Telecom.*, pp. 3-4, 1993.

[12] L. Cellario, D. Sereno, M. Giani, P. Blocher and K. Hellwig, "A VR-CELP Codec Implementation for CDMA Mobile Communications," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 1, pp. 281-284, 1994.

[13] K. Kikuiri, N. Naka and T. Ohya, "Super-Frame Based Source Controled Variable Rate Coding Using Approximated Trellis Diagram," *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 2, pp. 185-188, 2003.

[14] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier Science B.V., Amsterdam, 1995.